

A Queueing Characterization of Information Transmission over Block Fading Rayleigh Channels in the Low SNR Regime

Yunquan Dong, Pingyi Fan, *Senior Member, IEEE*

{dongyq08@mails, fpy@mail}.tsinghua.edu.cn

Department of Electronic Engineering, Tsinghua University, Beijing, China;

Abstract—Unlike the AWGN (additive white gaussian noise) channel, fading channels suffer from random channel gains besides the additive Gaussian noise. As a result, the instantaneous channel capacity varies randomly along time, which makes it insufficient to characterize the transmission capability of a fading channel using data rate only. In this paper, the transmission capability of a buffer-aided block Rayleigh fading channel is examined by a constant rate input data stream, and reflected by several parameters such as the average queue length, stationary queue length distribution, packet delay and overflow probability. Both infinite-buffer model and finite-buffer model are considered. Taking advantage of the memoryless property of the service provided by the channel in each block in the low SNR (signal-to-noise ratio) regime, the information transmission over the channel is formulated as a *discrete time discrete state D/G/1* queueing problem. The obtained results show that block fading channels are unable to support a data rate close to their ergodic capacity, no matter how long the buffer is, even seen from the application layer. For the finite-buffer model, the overflow probability is derived with explicit expression, and is shown to decrease exponentially when buffer size is increased, even when the buffer size is very small.

Index Terms—block Rayleigh fading channel, channels service, buffer-aided communications, queueing analysis, queue length distribution, packet delay, overflow probability.

I. INTRODUCTION

In the pioneering work of Shannon in 1948 [1], the capacity of a band limited AWGN (additive white Gaussian noise) channel with an average transmit power constraint was given, which specifies the fundamental transmission limit of AWGN channels,

$$C = W \ln \left(1 + \frac{P}{N} \right), \quad (1)$$

where W is the limited bandwidth and P/N is the SNR (signal-to-noise ratio).

This means that by sufficiently involved encoding systems, although the transmitted signal is perturbed by the random noise, we can still transmit digits at the rate of C nats per second for the given SNR, with arbitrarily small frequency of errors. Later, many capacity approaching coding schemes such as Turbo codes [2] and LDPC codes [3] were developed. In this way, the channels (especially wired ones) can be abstracted into a data pipeline, which delivers data from one place to another at a constant rate, regardless of its contents. Therefore, AWGN channels can be described by the single parameter of data rate. This separability of data rate and other parameters such as delay and overflow probability is also the basis of the OSI (Open Systems Interconnection) model [4], [5], which is widely used in internet/telecommunication systems.

However, due to the fading property of wireless channels, this is very different for wireless communications. In fact, the communication capability of fading channels has also been investigated for more than 70 years, from physical layer to cross-layer ways.

A. Physical layer evaluations of fading channels: Capacity

Usually, wireless communications suffer from random channel gains besides the additive Gaussian noise. Therefore, the corresponding instantaneous capacity is also a random variable. Furthermore, the wireless channel capacity also depends on what is known about the channel gain h (CSI, channel state information) at the transmitter and receiver. Particularly, the coherent channel model assumes perfect CSI. On the contrary, the noncoherent model not only assumes no CSI available, but also implicitly ignores the correlation between the fading coefficients over time, since the effect of this correlation will be dwindled by the asymptotic analysis at low SNR (signal-to-noise) [6]. For the coherent channel, transmitter and receiver CSI allows the transmitter to adjust both its power and rate according to the time varying channel gain. In this case, the *ergodic capacity* is defined as

$$C = I(X; Y, S) = \max_{P(h): \int P(h)p(h)dh = \bar{P}} \mathbb{E}_h W \ln \left(1 + \frac{h^2 P(h)}{N} \right) \quad (2)$$

where \bar{P} is the average power, $P(h)$ is the power allocation policy, $p(h)$ is the channel gain distribution and \mathbb{E}_h is the expectation operator on h . According to [10], the optimal solution is the water-filling power allocation over channel states.

When CSI is only available at the receiver, the transmitter can not perform power adaption. In this case, the transmitted signal is independent from channel gain and the ergodic capacity becomes

$$\begin{aligned} C &= I(X; Y, S) = I(X; S) + I(X; Y|S) = I(X; Y|S) \\ &= \mathbb{E}_h W \ln \left(1 + \frac{h^2 \bar{P}}{N} \right). \end{aligned} \quad (3)$$

Besides, its channel capacity can only be given by a multi-character definition as [7], [8]

$$C = \lim_{n \rightarrow \infty} \frac{1}{n} \sup_{p(x)} I(X^n; Y^n), \quad (4)$$

where $p(x)$ is the distribution of the input to the transmitter.

On the one hand, in order to approach ergodic capacity, the code lengths must be long enough to average over the fading

process, which is difficult to implement in practice. On the other hand, ergodic capacity is the inherent characteristic of the channel and helps understand the performance limits of wireless communications.

However, a transmission strategy for ergodic capacity must incorporate every state of the fading process. Thus those very poor channel states will typically reduce ergodic capacity greatly. To this end, an alternative capacity definition for fading channels with receiver CSI is the *outage capacity*, which specifies the maximum mutual information rate that can be transmitted in every channel realization except a subset whose probability is less than ϵ [10], [11], [12],

$$C = \max_R \{ R : \Pr \{ W \ln(1 + \bar{P}/N) < R \} < \epsilon \}. \quad (5)$$

It is clear that ϵ -outage capacity considers mainly on the worst cases and on the reliability in the communication. So there are wastes of channel capability in those very good channel conditions.

As the extension of the AWGN channel to fading channels, the concepts as listed above present some descriptions of fading channels using average data rate from the physical layer.

However, wireless networks have been widely deployed for a variety of purposes, where QoS (quality of service) sensitive applications are increasingly of interest. Each application has its own specified requirement on throughput, reliability, delay, and delivery ratio. Therefore, only using a long term throughput like ergodic capacity/outage capacity to characterize fading channels is not enough.

In this paper, we consider such a problem of *what kind of service fading channels can provide* and will characterize the transmission capability of a buffer-aided block Rayleigh fading channel from multiple aspects such as data rate, packet delay, stationary queue length distribution and overflow probability.

B. Cross layer evaluations: EC and queueing formulations

High demands on the QoS, especially when packet delay is the main concern, have fueled substantial research interest in modeling fading channels from upper layers. In [13], the statistics of the packet dropping process due to buffer overflow was investigated. Overflow probability and delay were also studied in [14], where data arrives in a on-off process and suffers from independent Bernoulli channel errors. The tail distribution of packet delay was studied under both round-robin scheduling and multiuser diversity scheduling algorithms in [15]. Besides, cross-layer joint-adaptation mechanisms which optimize the physical-layer adaptive modulation and coding and the link-layer packet fragmentation was proposed in [16]. In the paper, more realistic correlated fading channels were discussed using the finite state Markov Chain models (FSMC).

Besides above performance analysis, more fundamental investigations on the transmitting limits of fading channels are also needed from a cross layer perspective. In fact, in face of time varying property of fading channels and the booming variety of traffics in the system, physical layer capacity is no longer a sufficient characterization of the transmission capability of fading channels. On the contrary, the problem

what kind of service fading channels can provide can only be answered in a cross layer way.

Particularly, the authors in [17] proposed a link layer channel model termed effective capacity (EC) as a unified framework to consider parameters such as traffic rate and packet delay bound violations. Using the moment generating method and large deviation theory, two EC functions were established, namely the non-empty buffer probability $\gamma^{(c)}(\mu)$ and decaying exponent $\theta^{(c)}(\mu)$. To summarize, the EC link model aims to characterize wireless channels in terms of functions that can be easily mapped to link-level QoS metrics. However, this channel model involves a lot of inevitable approximation, which restricted its application.

On the other hand, wireless communications and queueing theory are connected with each other naturally. Firstly, due to the fluctuation of the instantaneous channel capacity, a buffer must be used at the transmitter to match the source traffic with the channel transmission capability in the engineering practice, which is a typical queueing problem. In theory analysis, Gallager has used queueing theory in the collision resolution problem in multi-access channels [18] in the mid 1980s. Later, in his work with Telatar [19], they used tools from both queueing theory and information theory to exhibit aspects from collision resolution approach and capacity bounds of a multi-access problem. More recently, Gallager and Berry discussed the optimal power allocation policy under some delay constraints in a finite-buffer aided point-to-point wireless communication system [20]. Particularly, two important questions were asked in the conclusion part of the paper. First, due to the fading character of wireless channels, the transmission rate of fading channels depends greatly on the link layer power allocation and rate allocation schemes, which makes the boundary between physical layer and link layer unclear. Therefore, *whether the layered communication protocol design in wired-line communications is reasonable for the wireless situations* is an important problem. For this question, it has been partially answered in [21], where it is proved that *i.i.d.* fading channels can support a constant rate data stream, just like those wired communication links. Therefore, the separation between physical layer and network layer is reasonable in the *i.i.d.* situations. The second problem is *what is the overflow probability of a finite buffer aided wireless link when the arrival rate is constant*. In this paper, this question will be answered for the case of block rayleigh fading channels in the low SNR regime.

In brief, to characterize what kind of service fading channels can provide in a cross-layer way, the channel can be examined by a constant rate data input to the buffer, and described from multi aspects such as queue length distribution, packet delay, and overflow probability.

However, there are some challenges in applying queueing theory to this problem. For a block fading channel, its channel gain varies block by block. Particularly, its channel gain is a positive real number ranging from zero to infinity. Therefore, the input/output process has to be modeled as a *discrete time continuous state* Markov process, for which few references exist. So one has to resort to other techniques such as stochastic process as in [22] or transform the continuous state

space to discrete state space by quantization [23].

C. Overview of this paper

This paper focuses on characterizing the transmission capability of block fading Rayleigh channels in the low SNR regime, in a cross-layer way.

Although insufficient, previous results in the framework of information theory still serves as an irreplaceable guidance for this work. Therefore, transmission rate is one of the most important parameter for fading channels. In addition, several QoS parameters are evaluated at the same time.

On the other hand, related cross-layer analysis and studies are usually asymptotic results when buffer size is very large. Therefore, this paper focuses on more detailed and explicit characterizations of fading channel transmission capability.

Specifically, the transmission capability of a fading channels is examined by a constant rate input data stream, and represented in terms of packet delay, queue length distribution and overflow probability.

It is assumed that data arrives at the buffer in packets. Although it is well known that it can be modeled as a Markov process and the overflow probability decreases exponentially, the following challenges exist:

- 1) The buffer-aided transmission over fading channels corresponds to a *discrete-time continuous-state* Markov process, since the channel gain is continuously distributed while the transmission over the time is in blocks, on which little is known.
- 2) Known results on the overflow probability are asymptotic ones by assuming the buffer size to be very large, explicit expressions are non-available either.

However, in this paper, the data transmission over a block fading channel is formulated successfully as a *discrete-time and discrete-state* queueing problem (Markov Chain), which applies to both small and large buffers. The key idea behind this formulation is the *memoryless property of exponential distributions*. In fact, the service provided by a Rayleigh fading channel in one block is negative exponentially distributed in the low SNR regime. In this situation, although part of the service capability of the block has been consumed by a previous packet (or a part of the packet), the remaining service capability of this block still follows the same distribution as itself, which can be seen as the service provided by a brand new block. In this way, the service time of a packet can be seen as the integer part of its actual service time. Thus, the state space, i.e., the queue length at discrete epochs (the beginning of each block) will be discrete. That is, the original discrete time continuous state queueing process is transformed in to a more trackable discrete time discrete state Markov chain.

With respect to the existing literature, the main contributions of this paper can be summarized as follows.

- a) A trackable Markov Chain was established to model the information transmission over block Rayleigh fading channels in the low SNR regime.
- b) For the infinite-buffer model, the stationary distributions was obtained. Closed form of average packet delay was presented, as a function of input data rate.

- c) For the finite-buffer model, the stationary queue length distribution and the average packet delay were derived.
- d) For a given buffer size, the overflow probability is given in an explicit expression, no matter the buffer size is small or large.

The rest of this paper is organized as follows. The communication system model is presented in Section II and the corresponding queueing model is formulated in Section III. After that, the infinite-buffer model is studied, where the stationary queue length distribution and packet delay is obtained in Section IV. In Section V, the finite-buffer model is investigated and its stationary queue length distribution, packet delay and overflow probability are derived on the basis of results on infinite-buffer model. The obtained result will also be presented via numerical results in Section VI. Finally, we will conclude our work in section VII.

II. SYSTEM MODEL

Consider a point-to-point communication over a block-fading Rayleigh channel with additive white Gaussian noise [24]. In such channels, the channel gain is modeled by a Rayleigh distributed random variable, which is stationary and ergodic. The channel gain stays fixed over each block of a certain amount of channel uses, and varies independently in different blocks. Let T_B be the block length, h_n be the time varying channel gain during the n -th block and $\gamma_n = h_n^2$ be the corresponding power gain. Then h_n is Rayleigh distributed and γ_n is exponentially distributed, whose probability density functions (*p.d.f.*) are given by, respectively,

$$p_h(x) = \frac{1}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} \quad \text{and} \quad p_\gamma(x) = \frac{1}{2\sigma^2} e^{-\frac{x}{2\sigma^2}}, \quad (6)$$

for $x > 0$, where h_n is independent of any other $h_{n'}$ for $n \neq n'$, and so is γ_n . In fact, $h = \sqrt{r_I^2 + r_Q^2}$, where r_I and r_Q are independent Gaussian random variables with zero mean and variance σ^2 .

The transmission model considered in this paper follows the model given in [20], as shown in Fig. 1. Assume that the higher layer application maintains a data stream with constant rate R . Since the service provided in each block is a random variable, we have to adjust the actual transmission rate over the fading channel according to the channel states. Therefore, a First In First Out (FIFO) buffer is used at the transmitter side to match the source traffic stream with the channel service in each block. Assume that the data are served by packets and the packet size equals to the traffic arriving at the buffer in each block, i.e., $L_p = RT_B$. Let $Q(n)$ be the queue length of the buffer in *packets* at the start of block n . Note that $Q(n)$ is not necessarily an integer at discrete times of n^+ . We use $D(n)$ to denote the number of blocks that the packet arriving in block n will spend in the queue.

Let P denote the transmit power, W is the transmit bandwidth and N_0 is the noise power spectral density. Let α and d denote the path loss exponent and the distance between the transmitter and receiver, respectively. Then the instantaneous capacity of the block Rayleigh fading channel in *nats* is

$$c_n = W \ln \left(1 + \frac{\gamma_n P d^{-\alpha}}{W N_0} \right), \quad (7)$$

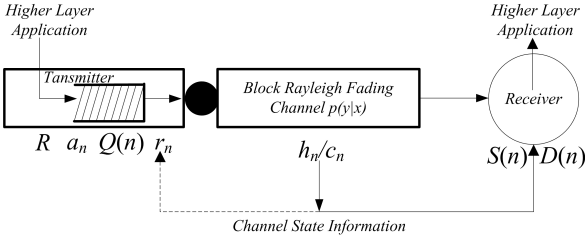


Fig. 1. The transmission system model

and the service provided by the fading channel in one block is $s_n = c_n T_B$.

Following the concept of channel service in [21], which combines dimensions of data rate and time, define S_k as the total amount of service provided by the fading channel in k successive blocks, we have

$$S_k = \sum_{m=1}^k s_m. \quad (8)$$

The average received SNR is defined as $\rho = \frac{2\sigma^2 P}{WN_0 d^\alpha}$.

III. QUEUEING MODEL FORMULATION: INFINITE BUFFER SIZE CASE

As is shown, the channel service in each block is a random variable of positive real numbers. Therefore, the queue size $Q(n)$ can not be guaranteed an integer when counting at some fixed epochs such as the beginning of each blocks n^+ . Therefore, the queueing process is a *discrete time continuous state* Markov process. Unfortunately, few results are available for such processes. However, some transformations are performed on the queueing process in this section, and finally, we will construct a discrete time discrete state $D/G/1$ queue when the average SNR is low. Particularly, the buffer size is assumed to be infinite long.

A. The service time of each packet

Firstly, the cumulative distribution function (CDF) of the instantaneous capacity c_n is given by

$$F_c(x) = 1 - e^{-\frac{1}{\rho} \left(e^{\frac{x}{WT_B \rho}} - 1 \right)}. \quad (9)$$

In the low SNR scenario considered in this paper, especially for the wide band communications, we have the following equation holds

$$F_c(x) = 1 - e^{-\frac{x}{WT_B \rho}}. \quad (10)$$

Remark 1: It should be noted that, if the average SNR is not so small, or equivalently the transmit power is not very small, the *p.d.f.* of channel service in each block is no longer negative exponential distribution and the approximation (10) of (9) will not be accurate anymore, as shown by the example in Fig. 2. Therefore, in the case of middle/high SNR or non-Rayleigh fading analysis, the memoryless property of the service provided in each block won't hold any more. As a result, we have to resort to other techniques such as state space quantization or stochastic analysis in [23].

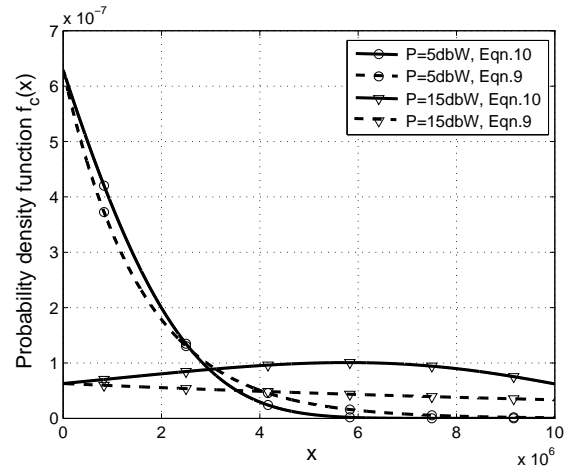


Fig. 2. The approximation of the *p.d.f.* of channel service of each block for middle SNR case

Therefore, in the low SNR regime, the CDF of the service in one block (s_n) will be given by

$$F_s(x) = 1 - e^{-\frac{x}{\nu}}, \quad (11)$$

which is an negative exponential distribution, where ν is defined as

$$\nu = WT_B \rho. \quad (12)$$

It can be seen that the service of k successive blocks (S_k) follows the Gamma distribution, whose *p.d.f.* is given by

$$f_S(x) = \frac{1}{\Gamma(k) \nu^k} x^{k-1} e^{-\frac{x}{\nu}}, \quad (13)$$

where $\Gamma(k) = \int_0^\infty e^{-t} t^{k-1} dt$ is the Gamma function.

Definition 1: The *service time* T_n of a packet is the number of *complete blocks* of the period in which the service of the packet is finished.

Therefore, T_n is a non-negative integer random variable. Firstly, the probability that a packet is served within one block can be derived as follows,

$$p_0 = \Pr\{T_n = 0\} = \Pr\{s_n > L_p\} = e^{-\frac{L_p}{\nu}} \stackrel{(\Delta)}{=} e^{-\theta}, \quad (14)$$

where $\theta = \frac{L_p}{\nu}$.

Remark 2: We say that the service time of a packet is zero if it is served within one block, for which the detailed explanations are listed as follows. For the block in which the packet is served, although part of its service capability has been consumed, the remaining amount of service follows the same probability distribution as that provided by a complete block, according to the memoryless property of negative exponential distribution. So it can continue to serve the next packet like another complete block. In this sense, the service time of this packet is zero.

Remark 3: Besides, the ergodic capacity reduces to $C_e = W\rho$ in the low SNR regime. So we have $\theta = \frac{L_p}{\nu} = \frac{tT_B}{WT_B \rho} = \frac{R}{C_e}$. Thus, θ is the ratio between traffic rate and ergodic capacity. Secondly, from the queueing aspect, ν is the average amount of service provided in each block. Then θ is the average number of blocks needed for one packet, which is

also the average service completion interval. Finally, θ can further be interpreted as $\frac{L_p}{\nu}/1$, which is the ratio between the average service time and the average arrival interval, which is often referred as to the *traffic load*.

Similarly, with $F_s(x)$, $f_{s_k}(x)$ and recall that $S_{k+1} = S_k + s_{k+1}$, the probability that the service time of a packet is k ($k \geq 1$) blocks can be obtained as

$$\begin{aligned} p_k &= \Pr\{T = k\} \stackrel{(a)}{=} \Pr\{S_k \leq L_p, S_{k+1} > L_p\} \\ &\stackrel{(b)}{=} \int_0^{L_p} f_{s_k}(x_k) dx_k \int_{L_p - x_k}^{\infty} f_{s_{k+1}}(s_{k+1}) ds_{k+1} \\ &= \frac{1}{k!} e^{-\frac{L_p}{\nu}} \left(\frac{L_p}{\nu}\right)^k = \frac{1}{k!} e^{-\theta} \theta^k, \end{aligned} \quad (15)$$

where (b) holds because S_k and s_{k+1} are independent from each other. Most importantly, (a) is assumed to hold, also in the sense of the memoryless property of negative exponential distributions, which is formally given as follows.

If X is an negative exponentially distributed random variable, then X is memoryless, namely,

$$\Pr\{X \leq s + t | X > s\} = \Pr\{X \leq t\}. \quad (16)$$

Therefore, when we say that the service time of a packet is $T_n = k$, the packet is actually completed in $k + 1$ blocks, as shown in (15.a). However, only a part of the service capability of block $n+1$ is used. So it can continue its service for the next packet in the queue. As is shown by (11), the channel service of each block in the low SNR regime follows the negative exponential distribution, which is memoryless. Therefore, the remaining service capability of the $n+1$ -th block can be seen as the service that can be provided by a new block. In this sense, the packet uses only k blocks. Similarly, if a packet is completed within one block as in (14), it is defined that its service time is zero, i.e., $T_n = 0$.

B. The Markov Chain model

In this way, the problem of information transmission over a block fading Rayleigh channel can be transformed into a classical discrete time $D/G/1$ queueing problem. Its arrival process is $\{A_n = 1, n \geq 1\}$ (unit: L_p) and its service time T_n (unit: block) is a Poisson distributed random variable, whose probability generating function (PGF) is given by

$$G(z) = \mathbb{E}[z^{T_n}] = e^{-\theta} + \sum_{k=1}^{\infty} \frac{1}{k!} e^{-\theta} (\theta z)^k = e^{\theta(z-1)}. \quad (17)$$

According to the relationship of the moments and PGFs, we have

$$\mathbb{E}[T_n] = G'(t)|_{t=0} = \theta. \quad (18)$$

Remark 4: The arrival process considered is constant and one packet arrives in each block. Then the expected number of packets arriving during a service time also equals to θ . Throughout the paper, it is assumed that $\theta < 1$ so that the queue is stable.

Up to now, the information transmission over the channel is formulated as a late arrival system queueing model with immediate access, in which each packet arrives at the end of

a block (k^-) and leaves at the beginning of the block (k^+). The service of each packet is also assumed to start at the beginning of a block. If the service time of a packet is zero, i.e., $T_n = 0$, it is assumed to leave the queue immediately at the beginning of the block following its arrival.

Let Q_n^+ be the number of packets in the queue at the time of n^+ . Particularly, n^+ may not necessarily a departure epoch of a packet. In this case, n^+ is actually lies in the service time of a certain packet. Then the remaining service time of the packet after n^+ is a random variable and will determine the value of Q_{n+1}^+ together with the value the channel service in block $n+1$. That is to say, the value of Q_{n+1}^+ is determined not only by Q_n^+ but also by how many of the current packet has been served before n^+ . Therefore, the arbitrary time queue length process $\{Q_n^+, n \geq 0\}$ is usually not a Markov chain.

Define τ_n be the departure epoch of packet n . Assume that the packet will leave the buffer at the same time as packet $n-1$ if its service time is zero, i.e., $T_n = 0$ and $\tau_n = \tau_{n-1}$. Therefore, τ_n^+ is the aftereffectless time of the queue length process. Let $L_n^+ = Q(\tau_n^+)$ be the number of packets in the buffer after the departure of the n -th packet. Then $\{L_n^+, n \geq 1\}$ is a Markov chain, which is called the embedded Markov chain (EMC) of queue length process $\{Q_n^+, n \geq 0\}$. Since the arrival process is $\{A_k = 1, k \geq 1\}$, i.e., one packet each block, there will be T_n packet arriving in the service time of the n -th packet T_n . Then we have

$$L_{n+1}^+ = \begin{cases} L_n^+ - 1 + T_n, & L_n^+ \geq 1 \\ T_n, & L_n^+ = 0. \end{cases} \quad (19)$$

The transition probability matrix of $\{L_n^+, n \geq 1\}$ will be given by

$$\mathbf{P} = \begin{bmatrix} p_0 & p_1 & p_2 & p_2 & \cdots \\ p_0 & p_1 & p_2 & p_2 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & p_0 & p_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (20)$$

The details of packets arrival and departure of some elements ($p_{00}, p_{01}, p_{10}, p_{11}$) in matrix \mathbf{P} are illustrated in Fig. 3 as some examples.

Discussion 1: Poisson Arrivals.

Suppose the arrival process is a Poisson process with intensity λ rather than constant process. Then the number of packets arriving at the buffer in each block will be a random variable of non-negative integer, i.e., $A_k^\lambda \in (-\infty, 0)$. Its PGF is given by

$$G_A^\lambda(z) = \mathbb{E}[z^{A_k^\lambda}] = e^{\lambda(z-1)}. \quad (21)$$

Assume the buffer is empty at the departure of n -th packet, the probability that the buffer is still empty at the departure of the $n+1$ -th packet will be

$$q_0 = \sum_{k=0}^{\infty} \Pr\{T_n = k\} \Pr\left\{\sum_{j=1}^k A_j^\lambda = 0\right\} = e^{\theta(e^{-\lambda}-1)}. \quad (22)$$

Similarly, the probability that the queue size is i at the

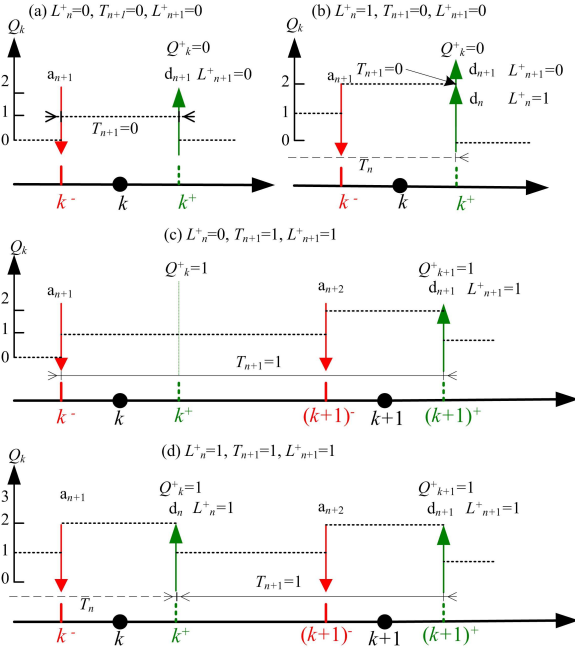


Fig. 3. Examples for the queueing model

departure of the $n + 1$ -th packet will be

$$q_i = \sum_{k=1}^{\infty} \Pr\{T_n = k\} \Pr\left\{\sum_{j=1}^k A_j^\lambda = i\right\}. \quad (23)$$

Denote $A_{T_n}^\lambda$ be the number of packets arriving in T_n blocks, we know that it is a compound Poisson process, whose PGF is given by

$$G_{A_{T_n}^\lambda} = e^{\theta(G_A^\lambda - 1)} = e^{\theta(e^{\lambda(z-1)} - 1)} \quad (24)$$

Therefore, one has

$$q_i = \mathbb{E}_{T_n} [\Pr\{A_{T_n}^\lambda = i\}] = \frac{G_{A_{T_n}^\lambda}^{(i)}(0)}{i!}. \quad (25)$$

In this way, the transition probability matrix of the queue size at departure epochs when arrival process is Poisson is given by:

$$\mathbf{P}^\lambda = \begin{bmatrix} q_0 & q_1 & q_2 & q_2 & \cdots \\ q_0 & q_1 & q_2 & q_2 & \cdots \\ 0 & q_0 & q_1 & q_2 & \cdots \\ 0 & 0 & q_0 & q_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (26)$$

The following analysis is focused on the case when the arrival process is constant. Results for Poisson arrival can be obtained by simply replacing \mathbf{P} with \mathbf{P}^λ and are omitted here.

IV. QUEUE LENGTH DISTRIBUTION AND DELAY ANALYSIS

In this section, we will perform a detailed investigation on the queue length process in terms of stationary queue length distribution and packet delay.

A. The Stationary Queue Length at the Departure Epochs

Let $L^+ = \lim_{n \rightarrow \infty} L_n^+$ be the limit of the queue length process, which is called the *stationary queue length at departure epochs*. Denote

$$\pi_j = \Pr\{L^+ = j\} = \lim_{n \rightarrow \infty} \Pr\{L_n^+ = j\}, \quad j \geq 0, \quad (27)$$

then the vector $\vec{\pi} = \{\pi_0, \pi_1, \dots\}$ is the stationary queue length at the departure epochs.

Theorem 1: If $\theta < 1$, the PGF of the stationary queue length at the departure epochs is given by:

$$L^+(z) = \frac{(1-\theta)(1-z)}{1 - ze^{\theta(1-z)}}, \quad (28)$$

where $\theta = \frac{L_p}{\nu}$.

With $L^+(z)$, the average queue length can be obtained by

$$\mathbb{E}[L^+] = \lim_{z \rightarrow 1} L^{+'}(z) = \frac{\theta(2-\theta)}{2(1-\theta)}. \quad (29)$$

The variance of queue length can also be obtained as

$$\begin{aligned} \text{Var}[L^+] &= \lim_{z \rightarrow 1} \left[L^{+''}(z) + L^+(z)' - (L^+(z)')^2 \right] \\ &= \frac{12\theta - 18\theta^2 + 10\theta^3 - \theta^4}{12(1-\theta)^2}. \end{aligned} \quad (30)$$

While the average queue length determines the buffer size, the standard deviation $\sigma_{L^+} = \sqrt{\text{Var}[L^+]}$ is an indicator of the “spread” of the queue length distribution. For a smaller σ_{L^+} , the queue length tends to be very close to the mean value. For large standard deviations, the queue length is spread out over a wider range of values.

It is known that there is a one-to-one correspondence between the probability distribution and its PGF. Therefore, the distribution of the queueing process at departure epochs is totally determined by L^+ , which will be derived later.

Besides, by a similar discussion in [26] on the relationship between the stationary distribution at departure epochs and arbitrary epochs, we have the following proposition. Interested readers can refer [27] for more details, as well as the proof of *Theorem 1*.

Proposition 1: The stationary queue length distribution at departure epochs and arbitrary epochs are the same.

Therefore, we will have the stationary distribution of the queue length process at arbitrary epochs by *Theorem 1* and *Proposition 1*.

B. Length of Busy Period

The *busy period* B is defined as a time interval during which the channel continuously transmitting packets and an *idle period* I is a time interval when the buffer is empty. Thus the system repeats cycles of busy and idle periods.

Proposition 2: The PGF and average value of busy period B are given by, respectively

$$B(z) = \frac{-1}{\theta z} W(-\theta z e^{-\theta}), \quad \text{and} \quad \mathbb{E}[B] = \frac{\theta}{1-\theta}, \quad (31)$$

where $W(z)$ is the Lambert W function [28]. The average idle period is given by $\mathbb{E}[I] = \frac{1}{e^\theta - 1}$.

Proof: Note that the length of a busy period and the number of packets served during a busy period do not depend on the order of service. Thus, we can assume the Last Come First Service (LCFS) discipline. Then there is a recursive relationship in the busy period. That is, the busy period is composed of the service time of the first packet in the busy period, and all the periods generated by packets that arrive during the service time of the first packet. Since one packet arrives in each block, the number of packets arriving in the first service period equals to the length of the service period. We have

$$B = T + B^1 + B^2 + \cdots + B^T, \quad (32)$$

where $\{B^n, n = 1, 2, \dots, T\}$ are mutually independent random variables with the same distribution as B , which leads to

$$\mathbb{E}[z^B | T = k] = \mathbb{E}[z^{k+B^1+\cdots+B^k}] = [zB(z)]^k. \quad (33)$$

According to the whole probability formula, we have

$$\begin{aligned} B(z) &= \sum_{k=0}^{\infty} \Pr\{T = k\} \mathbb{E}[z^B | T = k] \\ &= \sum_{k=0}^{\infty} \Pr\{T = k\} [zB(z)]^k \\ &= G[zB(z)], \end{aligned} \quad (34)$$

where $G(z)$ is the PGF of the service time and given in (17).

Solving $B(z)$ from (34), we have $B(z) = \frac{1}{\theta z} W(-\theta z e^{-\theta})$ for $|z| < 1$.

Then the expected value of busys period can be derived by

$$\begin{aligned} \mathbb{E}[B] &= B'(z)|_{z=1} \\ &= G'[zB(z)][B(z) + zB'(z)]|_{z=1} \\ &= G'[1][1 + B'(1)] \\ &= \theta(1 + \mathbb{E}[B]), \end{aligned} \quad (35)$$

and we get $\mathbb{E}[B] = \frac{\theta}{1-\theta}$.

During the idle period, the buffer is continuously empty. Therefore, packets arriving in this period are all served within one block, namely,

$$\begin{aligned} \Pr\{I = k\} &= \Pr\{s_1 > L_p, \dots, s_k > L_p, s_{k+1} < L_p\} \\ &= p_0^k (1 - p_0); \end{aligned} \quad (36)$$

where $p_0 = \Pr\{s_n > L_p\} = e^{-\theta}$. It is clear that the idle period is a geometrically distributed random variable. Then the average idle period is given by

$$\mathbb{E}[I] = \sum_{k=1}^{\infty} k \Pr\{I = k\} = \frac{1}{e^{\theta} - 1}. \quad (37)$$

Because the average service time of a packet is $\mathbb{E}[T] = \theta$, the number of packets served in a busy period can be obtained as $\mathbb{E}[\Gamma] = \frac{\mathbb{E}[B]}{\mathbb{E}[T]} = \frac{1}{1-\theta}$.

It can be seen that $\mathbb{E}[B]$ is increasing while $\mathbb{E}[I]$ is decreasing with θ . The existence of $\mathbb{E}[B]$ and $\mathbb{E}[I]$ is a result of the fluctuation of the instantaneous channel capacity, which is the key difficulty of evaluating and using fading channels, and

is unavoidable. Particularly, larger $\mathbb{E}[B]$ means higher channel usage efficiency but larger packet delay. On the contrary, larger $\mathbb{E}[I]$ means wastes of channel and less delay. They can not be optimized at the same time. But some trade-offs can be made according to the specified application scenario.

C. Packet Delay

The *packet delay* D is defined as the time interval between the arrival of a packet and its departure. Firstly, the packet delay consists of a service time T . Secondly, if the packet arrives seeing a non-empty buffer, it must wait for a *waiting time* W for its service. Finally, for the formulation in this paper, there is another piece of time that the packet spends in the system, i.e., the *vestige time* V . In this paper, $T = k$ ($k = 0, 1, \dots$) means that the service of a packet is not finished until the $k+1$ -th block. According to the memoryless property of the negative exponential distribution, although part of the service ability has been consumed, it is considered as a brand new block. However, the packet still has to spend a part of that block in the system, which is called the *vestige time*. Thus, we know that

$$D = T + W + V. \quad (38)$$

Theorem 2: In the FIFO discipline, the average packet delay is given by

$$\mathbb{E}[D] = \frac{1}{2} + \theta + \frac{\theta^2}{2(1-\theta)} + \int_0^1 (x-1)e^{\frac{-\theta}{x}} dx. \quad (39)$$

Proof: By the formulation of the queueing model, it can be seen that the vestige time V is independent of service time T and waiting time W . Therefore, they can be obtained separately. See Appendix A for the details of the proof. ■

Theorem 3 shows the relationship between average packet delay and traffic load θ (equivalently the input data rate R). As known to all, ergodic capacity is the expected transmission rate averaged over all fading states, which is considered as the limiting transmission rate of a fading channel when physical layer coding delay is allowed to be large enough. Particularly, it was proved in [21] that, *i.i.d.* fading channels can actually support a constant data stream at the rate of ergodic capacity without queueing delay. However, it is seen from *Theorem 3* that as θ approaching 1, the average packet delay will go to infinity. This means that the block fading channel has difficulty in supporting a traffic rate near its ergodic capacity, even when the buffer size is infinite and considered from the application layer, and neglecting its physical layer coding delay.

V. QUEUEING MODEL FORMULATION: FINITE-BUFFER CASE

In the practical engineering, the size of a buffer must be finite. Thus it is more useful to investigate the finite buffer-aided communications.

Denote the buffer size as K . If a packet arrives at the buffer when the queue length is K , an overflow happens. In this section, we are interested in the overflow probability. In this case, the $(K+1) \times (K+1)$ transition probability matrix of

queue length process at the departure time $\{\widehat{L}_n^+, n \geq 1\}$ is given by

$$\widehat{\mathbf{P}} = \begin{bmatrix} p_0 & p_1 & p_2 & p_3 & \cdots & p_{k-1} & \widehat{p}_K \\ p_0 & p_1 & p_2 & p_3 & \cdots & p_{k-1} & \widehat{p}_K \\ 0 & p_0 & p_1 & p_2 & \cdots & p_{k-2} & \widehat{p}_{K-1} \\ 0 & 0 & p_0 & p_1 & \cdots & p_{k-3} & \widehat{p}_{K-2} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & p_1 & \widehat{p}_2 \\ 0 & 0 & 0 & 0 & \cdots & p_0 & \widehat{p}_1 \end{bmatrix} \quad (40)$$

where $\widehat{p}_k = \Pr\{T_n \geq k\} = 1 - \sum_{j=0}^k \frac{\theta^j}{j!} e^{-\theta}$ is the probability that the number of arrival packets (equal to the service time) is equal or more than k during a service of current packet.

A. Stationary Queue Length Distribution

Firstly, the stationary distribution of the queue length process for the infinite-buffer model can be obtained with its PGF (28) in *Theorem 1*. It is noted that $\pi_0 = 1 - \theta$.

Before the discussion on the overflow probability, let's introduce one lemma that will be used.

Lemma 1: Cauchy Integral Formula (extended) [29]

Let C be a simple closed positively oriented piecewise smooth curve on a domain D , and let the function $f(z)$ be analytic in a neighborhood of C and its interior. Then for every z_0 in the interior of C and every natural number n , we have that $f^{(n)}(z_0)$ is n -times differentiable at z_0 and its derivative is

$$f^{(n)}(z_0) = \frac{n!}{2\pi i} \oint_C \frac{f(z)}{(z - z_0)^{n+1}} dz, \quad n = 1, 2, \dots \quad (41)$$

Let C and C' are both circles centered on the origin with their radiuses $r < 1$ and $1 < r' < \frac{-1}{\theta} W_{-1}(-\theta e^{-\theta})$.

Since $L^+(z)$ is convergent within the unit circle, for $k \geq 1$, the stationary can be expressed by

$$\pi_k = \frac{1}{2\pi i} \oint_C L^+(z) \frac{dz}{z^{k+1}} = \frac{1-\theta}{2\pi i} \oint_{C'} \frac{1-z}{1-ze^{\theta(1-z)}} \frac{dz}{z^{k+1}}. \quad (42)$$

It can be seen that function $g(z) = \frac{1-z}{1-ze^{\theta(1-z)}} \frac{1}{z^{k+1}}$ has two singular points within circle C' , namely $z = 1$ and $z = \frac{-1}{\theta} W_{-1}(-\theta e^{-\theta})$, where $W_{-1}(z)$ is the lower branch Lambert W function. Particularly, $z = 1$ is one movable singularity since $\lim_{z \rightarrow 1} g(z) = 1$ is finite. Therefore, within circle C' , $g(z)$ can be considered as analytic. According to Cauchy-Goursat's theory, the integral of $g(z)$ along any closed curve in C' is a invariable. Then we have

$$\begin{aligned} \pi_k &= \frac{1-\theta}{2\pi i} \oint_{C'} \frac{1-z}{1-ze^{\theta(1-z)}} \frac{dz}{z^{k+1}} \\ &= \frac{1-\theta}{2\pi i} \oint_{C'} \frac{(z-1) \frac{e^{\theta(z-1)}}{z}}{\frac{e^{\theta(z-1)}}{z} - 1} \frac{dz}{z^{k+1}} \\ &\stackrel{(a)}{=} \frac{1-\theta}{2\pi i} \oint_{C'} (z-1) \frac{e^{\theta(z-1)}}{z} \sum_{j=0}^{\infty} \left(\frac{e^{\theta(z-1)}}{z} \right)^j \frac{dz}{z^{k+1}} \\ &= \frac{1-\theta}{2\pi i} \sum_{j=1}^{\infty} \oint_{C'} e^{j\theta(z-1)} \left(\frac{1}{z^{k+j}} - \frac{1}{z^{k+j+1}} \right) dz \\ &\stackrel{(b)}{=} (1-\theta) \sum_{j=1}^{\infty} \left[\frac{1}{(k+j-1)!} \left(e^{j\theta(z-1)} \right)^{(k+j-1)} \frac{1}{(k+j)!} \left(e^{j\theta(z-1)} \right)^{(k+j)} \right. \\ &\quad \left. - \frac{1}{(k+j)!} \left(e^{j\theta(z-1)} \right)^{(k+j)} \frac{1}{(k+j+1)!} \left(e^{j\theta(z-1)} \right)^{(k+j+1)} \right] \\ &= (1-\theta) \sum_{j=1}^{\infty} \left[\frac{1}{(k+j-1)!} (j\theta)^{k+j-1} - \frac{1}{(k+j)!} (j\theta)^{k+j} \right] e^{-j\theta} \end{aligned} \quad (43)$$

where (a) follows $\frac{1}{1-z} = \sum_{j=0}^{\infty} z^j$ for $|z| < 1$ and the fact that $\left| \frac{e^{\theta(z-1)}}{z} \right| < 1$ on circle C' , (b) follows *Lemma 1* and $e^{j\theta(z-1)}$ is analytic.

To insure the the following general item hold for all $k \geq 0$, it is defined that $\varphi_{-1} = 1$. Besides, define for $k \geq 0$

$$\varphi_k = (1-\theta) \sum_{j=1}^{\infty} \frac{1}{(k+j)!} (j\theta)^{k+j} e^{-j\theta} \quad (44)$$

Then we have

$$\pi_k = \varphi_{k-1} - \varphi_k. \quad (45)$$

Specifically, $\varphi_0 = \theta$. A simple proof is given as follows. Denote $x = -\theta e^{-\theta}$ for convenience. Remember the Lambert W function can be written as $W_0(x) = \sum_{n=0}^{+\infty} \frac{(-n)^{n-1}}{n!} x^n$ and $W'_0(x) = \frac{W_0(x)}{x(1+W_0(x))}$ holds [28]. Then we have

$$\begin{aligned} \varphi_0 &= (1-\theta) \sum_{j=1}^{\infty} \frac{1}{j!} (j\theta)^j e^{-j\theta} = (1-\theta) \sum_{j=1}^{\infty} \frac{(-jx)^j}{j!} \\ &= -x(1-\theta) \sum_{j=1}^{\infty} \frac{(-jx)^{j-1}}{j!} \cdot jx^{k-1} = -x(1-\theta) W'_0(x) \\ &= -x(1-\theta) \frac{W_0(x)}{x(1+W_0(x))} = \theta e^{-\theta} (1-\theta) \frac{-\theta}{-\theta e^{-\theta} (1-\theta)} \\ &= \theta. \end{aligned} \quad (46)$$

However, when we take the finite sum of (44) to approximate φ_0 , it needs more items for larger θ than smaller ones, as shown in Fig. 4.

On the other hand, $\Pr\{L^+ > K\}$ gives the probability that the queue length is over K in the infinite-buffer model, which can be seen as a *virtual overflow probability* of the infinite buffer model. Define the virtual overflow probability for a given K as

$$\begin{aligned} P_{overflow}^{\infty} &= \Pr\{L^+ > K\} = \sum_{j=K+1}^{\infty} \pi_j = \varphi_K \\ &= (1-\theta) \sum_{j=1}^{\infty} \frac{1}{(K+j)!} (j\theta)^{K+j} e^{-j\theta}. \end{aligned} \quad (47)$$

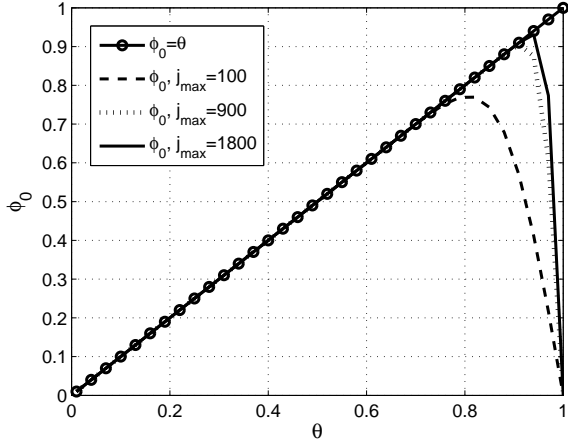


Fig. 4. The approximation of φ_0 by finite sums

Next, define $\vec{\pi}^K = \{\pi_0^K, \pi_1^K, \dots, \pi_K^K\}$ as the stationary distribution of the queue length process for the finite-buffer model, which is given by the following theorem.

Theorem 3: The stationary distribution of the queue length process for the finite-buffer model is given by

$$\begin{aligned} \pi_j^K &= \frac{\varphi_{j-1} - \varphi_j}{1 - \theta\varphi_K}, \quad j = 0, 1, \dots, K-1, \\ \pi_K^K &= \frac{\varphi_{K-1} - \theta\varphi_K}{1 - \theta\varphi_K}, \quad j = K, \end{aligned} \quad (48)$$

where $\varphi_{-1} = 1$ and φ_k is given by (44) for $k \geq 0$.

Proof: Define a *cycle* as the time interval between two consecutive service completions leaving an empty buffer behind. According to regenerative process theory, the statistical properties in one cycle is the same as that of the entire process when the process has been in stationary state.

For the convenience of the proof, let's give some definitions for the finite-buffer queue model, which are well defined under the assumption $\theta < 1$.

- N^K The number of blocks in one cycle.
- N_j^K The number of blocks in one cycle at whose beginning there are j packets in the buffer ($j = 0, 1, \dots, K$).
- S^K The number of packets served in one cycle.
- S_j^K The number of packets completions that there are j packets left in the buffer at its departure ($j = 0, 1, \dots, K-1$). It is clear that $S_0^K = 1$.
- μ_j^K The expected number of blocks until the next packets completion when a packet has been just completed with j packets left behind.

The corresponding quantities for the infinite-buffer model are denote by N , N_j , S , S_j and μ_j , respectively.

By the theory of regenerative process, we have $\pi_j^K = \frac{E[N_j^K]}{E[N^K]}$ and $\pi_j = \frac{E[N_j]}{E[N]}$.

Similarly, the transitions among states $\{0, 1, \dots, K-1\}$ are the same for both finite and infinite-buffer models. Besides, the service of each block is independent from each other. Thus the transitions among states above level K do not affect the number of downcrossing to any state $j \leq K-1$ during one

cycle in the infinite-buffer model. Note that this is not valid for $j = K$. Therefore,

$$E[N_j^K] = E[N_j], \quad j = 0, 1, \dots, K-1. \quad (49)$$

Let $\eta = \frac{E[N]}{E[N^K]}$, then we have

$$\pi_j^K = \eta\pi_j, \quad j = 0, 1, \dots, K-1. \quad (50)$$

On the other hand, we have $\mu_j^K = \mu_j$ for $0 \leq j \leq K-1$ and $\mu_j = \theta$ for $k \geq 1$. Particularly, for $j = 0$, there is some idle time besides the service time. We also have $E[S_j^K] = E[S_j]$ for $0 \leq j \leq K-1$. By the definition of N^K and N , their expected value are given by

$$\begin{aligned} E[N^K] &= \mu_0^K + \sum_{j=1}^{K-1} E[S_j^K] \mu_j^K, \\ E[N] &= \mu_0 + \sum_{j=1}^{\infty} E[S_j] \mu_j. \end{aligned} \quad (51)$$

Using the equations above, it follows that

$$E[N] = E[N^K] + \sum_{j=K}^{\infty} E[S_j] \theta. \quad (52)$$

According to the definition of η , we have $E[N]/\eta = E[N^K]$ holds. Dividing (52) by $E[N]/\eta$ and $E[N^K]$ on both sides, respectively, we have

$$\eta = 1 + \eta\theta \sum_{j=K}^{\infty} \frac{E[S_j]}{E[N]}. \quad (53)$$

By the theory of regenerative process, the statistical property of one cycle converges to that of the whole process and we have $\pi_j = \frac{E[S_j]}{E[S]}$. Since there is only one packet arrives in each block, it is clear that the length of a cycle equals to the number of packets served during the cycle. So we have $E[N] = E[S]$. Combining these discussions, we know that

$$\pi_j = \frac{E[S_j]}{E[N]}. \quad (54)$$

With this equation and solving η from (53) one can get

$$\eta = \frac{1}{1 - \theta \sum_{j=K}^{\infty} \pi_j} = \frac{1}{1 - \theta\varphi_K}. \quad (55)$$

Combining (45), (47), (50) and (55), *Theorem 3* is proved. ■

B. Packet Delay

Proposition 3: For the finite-buffer model and the FIFO discipline, the average packet delay (the period from the arrival of a packet to its departure) is given by

$$E[\hat{D}] = \frac{1}{2} + \frac{\sum_{j=0}^{K-1} \varphi_j - K\theta\varphi_K}{1 - \theta\varphi_K} + \int_0^1 (x+1)e^{-\frac{\theta}{x}} dx, \quad (56)$$

where $\varphi_{-1} = 1$ and φ_k is given by (44) for $k \geq 0$.

Proof: Since the stationary queue length distribution has been given in Theorem 3, the average queue length is given by

$$\begin{aligned}
 \mathbb{E}[\hat{L}^+] &= \sum_{j=0}^K j \pi_j^K \\
 &= \sum_{j=1}^{K-1} \frac{j(\varphi_{j-1} - \varphi_j)}{1 - \theta \varphi_K} + K \left(1 - \sum_{j=0}^{K-1} \frac{\varphi_{j-1} - \varphi_j}{1 - \theta \varphi_K} \right) \\
 &= \frac{1}{1 - \theta \varphi_K} \left(\sum_{j=0}^{K-1} \varphi_j - K \varphi_{K-1} \right) + K - \frac{K(1 - \varphi_{K-1})}{1 - \theta \varphi_K} \\
 &= \frac{\sum_{j=0}^{K-1} \varphi_j - K \theta \varphi_K}{1 - \theta \varphi_K}
 \end{aligned} \tag{57}$$

Since there is only one packet arrives in each block, the average time that a packet stays in the system (including its service time and waiting time) equals to the average queue length by the Little's law,

$$\mathbb{E}[\hat{T}_Q] = \mathbb{E}[\hat{L}^+]. \tag{58}$$

On the other hand, the vestige time of the finite-buffer model won't change much compared with the infinite-buffer model, especially for large K . Thus, the average vestige time is approximated by $\mathbb{E}[\hat{V}] = \mathbb{E}[V]$, which is given by (A.80).

Finally, the total average delay in the finite-buffer model is

$$\mathbb{E}[\hat{D}] = \mathbb{E}[\hat{T}_Q] + \mathbb{E}[\hat{V}] \tag{59}$$

which completes the proof. ■

C. Overflow probability

Define the overflow probability P_{overflow} of a finite size buffer as the long-run fraction of packets that are rejected due to the finite capacity of the buffer. This is also the probability that the buffer is in the overflow state at packet arriving epochs. Following the queueing analysis method, we will investigate the overflow probability of a finite size buffer in this section. Motivated by the result in [30], the overflow probability is obtained based on the results on the infinite-buffer model, and is given by the following theorem.

Theorem 4: Assume the buffer size is K (unit: L_p) and the overflow probability is P_{overflow} . For the queue model considered in this paper (one packet arrives each block), the overflow probability of the finite-buffer aided communication over block Rayleigh fading channel in the low SNR regime can be given by

$$P_{\text{overflow}} = \frac{(1 - \theta)\varphi_K}{1 - \theta\varphi_K}, \tag{60}$$

where $\theta = \frac{L_p}{\nu} = \frac{L_p}{WT_B\rho}$ and φ_K is given in (44).

Proof:

Under the assumption $\theta < 1$, both the finite and infinite-buffer models are stable, i.e., the queue will not be always in the overflow state, nor the queue length will go to infinity. In other words, each packet can be served within one block in the

average sense, which implies that the service time required by one packet equals to the service time offered by the channel in each block. As we know, the average service time required by each packet is θ , which is less than 1. On the other hand, the channel is busy except a probability of π_0 . So the service time provided by the channel in each block is $1 - \pi_0$ in the average sense. Then for the infinite-buffer model, the following equation holds.

$$\theta = 1 - \pi_0. \tag{61}$$

For the finite-buffer case, a fraction of P_{overflow} packets are discarded. Thus the actual amount of service time required in each block becomes $(1 - P_{\text{overflow}})\theta$. Therefore, we have

$$(1 - P_{\text{overflow}})\theta = 1 - \pi_0^K. \tag{62}$$

Combining (50), (61) and (62) and solving P_{overflow} , we have

$$P_{\text{overflow}} = \frac{(\eta - 1)(1 - \theta)}{\theta}. \tag{63}$$

By replacing the η with the result in (55), we finally get the overflow probability as

$$P_{\text{overflow}} = \frac{(1 - \theta)\varphi_K}{1 - \theta\varphi_K}, \tag{64}$$

which completes the proof. ■

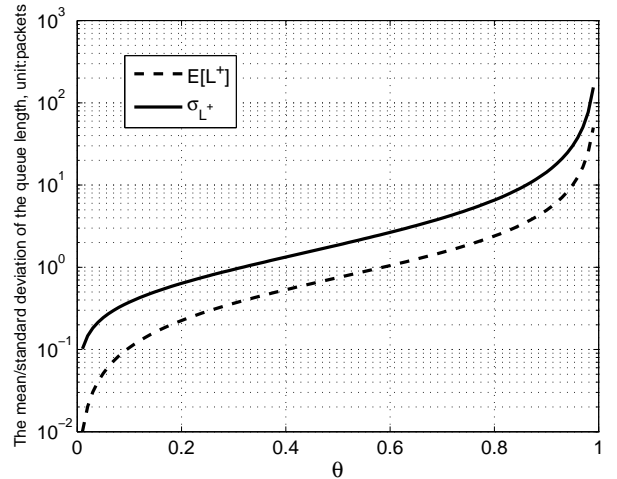


Fig. 5. The mean/standard deviation of the queue length in the infinite-buffer model

VI. NUMERICAL RESULTS

In this section, some numerical results are presented to illustrate the stationary distribution, packet delay, as well as the overflow probability for the finite-buffer model. The component variance is assumed as to be $\sigma^2 = 1$, the system bandwidth is $W = 5$ MHz and transmitting power is $P = -10$ dBW. Suppose that the distance between the transmitter and the receiver is $d = 1000$ m and the pathloss exponent is $\alpha = 4$. The block length is chosen as $T_B = 10^{-4}$ s.

According to its definition, we have $\theta = \frac{L_p}{WT_B\rho}$. Remember that $L_p = RT_B$ and the equivalent AWGN capacity of the

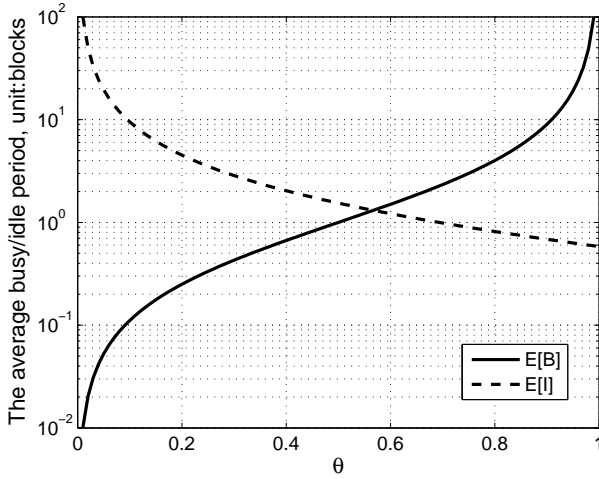


Fig. 6. The average busy/idle period of the infinite-buffer model

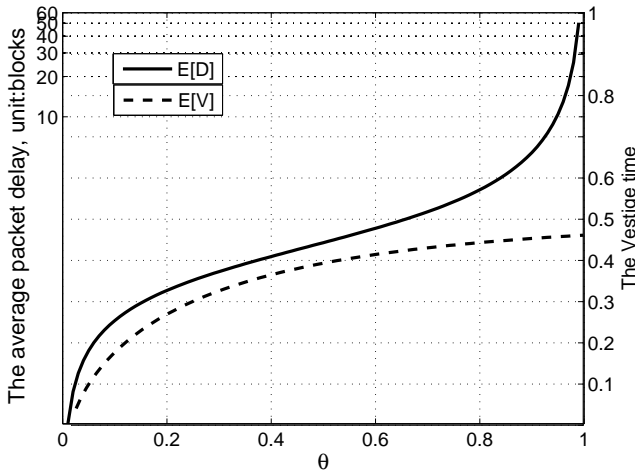


Fig. 7. The average packet delay of the infinite-buffer model. $E[D]$ corresponds to the log-scale y-axis on the left while $E[V]$ corresponds to the linear y-axis on the right.

channel is $C_a = W\rho$ in the low SNR regime. Therefore, θ can also be seen as the ratio between the traffic rate and the ergodic capacity.

For the infinite-buffer model, the mean value as well as the standard deviation of the queue length L^+ are presented in Fig. 5. When θ is increased, $E[L^+]$ and σ_L increases accordingly. Particularly, they will go to infinity when θ approaches 1. This means that, blocks fading channels can not support a data stream at a rate close to its ergodic capacity. Its busy period and idle period are given by (35) and shown in Fig. 6. It is seen that $E[B]$ increases while $E[I]$ decreases quickly with θ , which means that the utilization of channel is improved. However, the average packet delay will increase quickly at the same time, as shown in Fig. 7. Particularly, the average delay corresponds to the log-scale y-axis on the left and the average vestige time corresponds to the linear y-axis on the right, both in blocks. It is seen that the average vestige time $E[V]$ is also increasing with θ but never exceeds 0.5 blocks.

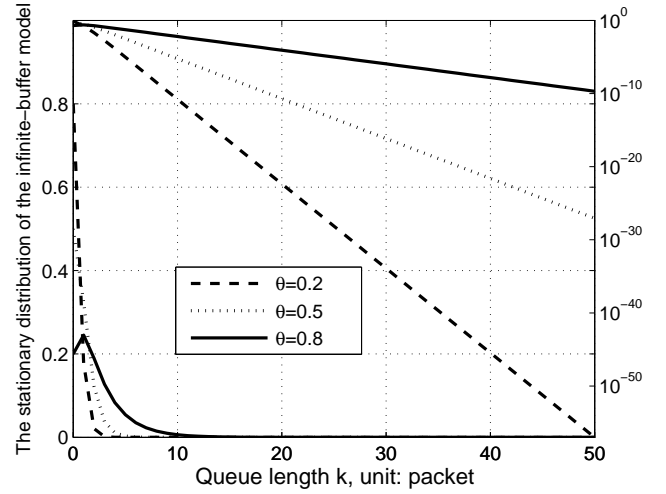


Fig. 8. The stationary queue length distribution of the infinite-buffer model. Each curve is presented both in linear y-axis on the left and log-scale y-axis on the right.

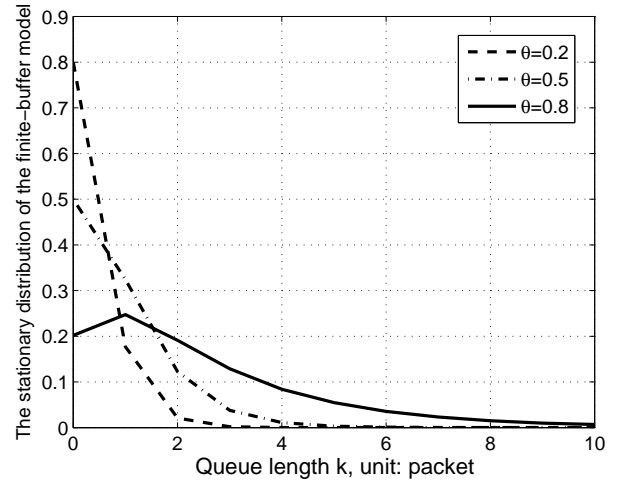


Fig. 9. The stationary queue length distribution of the finite-buffer model, buffer size $K = 10$ packets

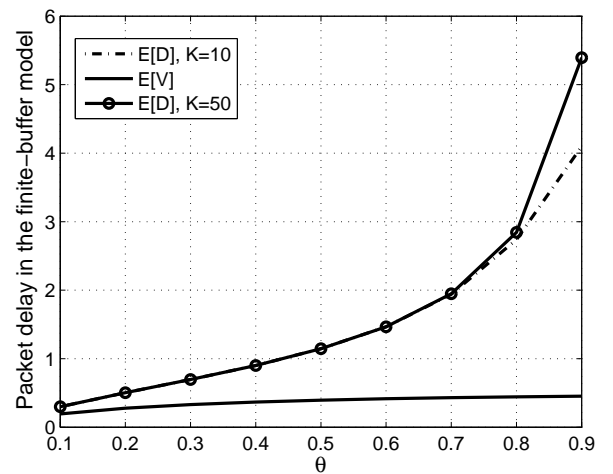


Fig. 10. The average packet delay (blocks) of the finite-buffer model

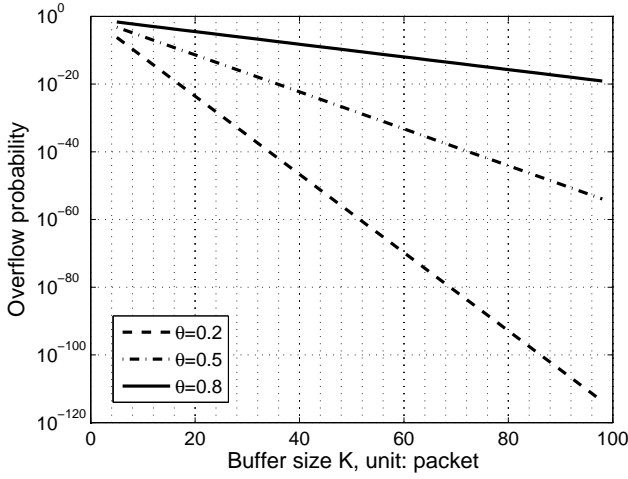


Fig. 11. The overflow probability versus buffer size (packets)

The stationary queue length distribution of the infinite-buffer model is shown in Fig. 8 for $\theta = 0.2, 0.5, 0.8$, where they are presented in both linear and log-scale y-axis. It is clear that the larger θ is, the slower π_k decreases. This is easy to understand because heavier traffic leads to longer queues. Most importantly, as the queue length grows, the probability that it occurs decreases approximately exponentially.

When the buffer size is finite, the stationary queue length distribution is slightly changed due to buffer overflows. Assume buffer size is $K = 10$ packets, the queue length distribution $\bar{\pi}^K$ is shown in Fig. 9. It is seen that the queue length is generally small and is increasing with θ . Similarly, the packet delay grows with both θ and buffer size, as shown in Fig. 10.

Finally, the overflow probability of a finite buffer is given in Fig. 11. For a smaller θ , i.e., slight traffics, the overflow probability is smaller and decreases more quickly. However, as is shown, all the overflow probability decrease exponentially in all cases.

VII. CONCLUSION

Modern communications requires the wireless channels to provides QoS (quality of service) guaranteed services. How to characterize and make full use of the service capability of fading channels is an urgent problem. In this paper, we solved the problem in the case of low SNR block Rayleigh fading channels by taking advantage of the memoryless property of their one block services. However, for general fading channels, the answers are quite unclear, which needs a lot of further efforts.

ACKNOWLEDGEMENT

This work was partly supported by the China Major State Basic Research Development Program (973 Program) No. 2012CB316100(2), National Natural Science Foundation of China (NSFC) No. 61171064 and NSFC No. 61021001.

APPENDIX A

PROOF OF THEOREM 3

Theorem 3: In the FIFO discipline, the average packet delay is given by

$$\mathbb{E}[D] = \frac{1}{2} + \theta + \frac{\theta^2}{2(1-\theta)} + \int_0^1 (x-1)e^{-\frac{\theta}{x}} dx. \quad (\text{A.65})$$

Proof:

A. The Service and Waiting Time

Since there is only one packet arrives in each block, the average time that a packet stays in the system ($S+W$) equals to the average queue length by the Little's law. So we have

$$\mathbb{E}[S+W] = \mathbb{E}[L^+] = \lim_{z \rightarrow 1} L^+(z) = \frac{\theta(2-\theta)}{2(1-\theta)}. \quad (\text{A.66})$$

B. The Vestige Time

To investigate the vestige time V , we have to define some new random variables. Firstly, for $k \geq 1$, define

$$\begin{aligned} U_k &= L_p - S_k, & U_k^+ &= U_k |_{U_k > 0}, \\ V_k &= \frac{U_k^+}{s_n}, & V_k^- &= V_k |_{V_k < 1}, \end{aligned} \quad (\text{A.67})$$

where s_n is the service provided by the channel in one block and S_k is the total amount of service provided by k successive blocks defined in (8). By the definition, U_k is the remaining part of a packet after k blocks of transmission, which is not necessarily positive. Using a positive condition, we get U_k^+ . In this case, there is really part of the packet untransmitted. Assuming the channel service of the next block is s_n , the remaining amount of data U_k^+ needs V_k blocks to be transmitted. However, V_k is the vestige time only if the remaining data can be transmitted within this block, which reduces to V_k^+ .

As one can see, $U_k \in (-\infty, L_p)$, $U_k^+ \in (0, L_p)$, $V_k \in (0, \infty)$ and $V_k^+ \in (0, 1)$. Particularly, the CDF of U_k can be derived as follows.

$$\begin{aligned} F_{U_k}(x) &= \Pr\{U_k \leq x\} = \Pr\{S_k \geq L_p - x\} \\ &= 1 - F_S(L_p - x) = \Gamma(k, \frac{L_p - x}{\nu}) \\ &= \frac{1}{\Gamma(k)} \int_{\frac{L_p - x}{\nu}}^{\infty} t^{k-1} e^{-t} dt, \end{aligned} \quad (\text{A.68})$$

where $\Gamma(k, x) = \frac{1}{\Gamma(k)} \int_x^\infty t^{k-1} e^{-t} dt$ is the upper Gamma function and $\nu = WT_B \rho$.

Therefore, the CDF of U^+ is

$$\begin{aligned} F_{U_k^+} &= \Pr\{U_k^+ \leq x | U_k^+ > 0\} = \frac{\Pr\{0 < U_k^+ \leq x\}}{\Pr\{U_k^+ > 0\}} \\ &= \frac{F_{U_k}(x) - F_{U_k}(0)}{1 - F_{U_k}(0)} = \frac{\int_{\frac{L_p-x}{\nu}}^\infty t^{k-1} e^{-t} dt - \Gamma(k, \theta)}{\gamma(k, \theta)}, \end{aligned} \quad (\text{A.69})$$

where $\gamma(k, x) = \frac{1}{\Gamma(k)} \int_0^x t^{k-1} e^{-t} dt$ is the lower Gamma function. For $x \in (0, L_p)$, taking derivative on $F_{U_k^+}$, we get the *p.d.f.* of U_k^+ as

$$f_{U_k^+}(x) = \frac{1}{\nu^k \gamma(k, \theta)} (L_p - x)^{k-1} e^{-\frac{L_p-x}{\nu}}. \quad (\text{A.70})$$

For some $x \in (0, +\infty)$, define D as the area in the $U_k^+ - s_n$ plane that $0 < U_k^+ < L_p$ and $s_n > \frac{U_k^+}{x}$. Since U_k^+ and the channel service of the next block (s_n) are random variables independent from each other, the CDF of V_k can be derived as follows.

$$\begin{aligned} F_{V_k}(x) &= \Pr\{V_k \leq x\} = \Pr\{U_k^+ \leq x s_n\} \\ &= \iint_D f_{U_k^+}(u) f_{s_n}(s) du ds \\ &= \int_0^{L_p} f_{U_k^+}(u) du \int_{\frac{u}{x}}^\infty \frac{1}{\nu} e^{-\frac{s}{\nu}} ds \\ &= \frac{1}{\nu^k \gamma(k, \theta)} \int_0^{L_p} (L_p - u)^{k-1} e^{-\frac{1}{\nu}(L_p - u + \frac{u}{x})} du. \end{aligned} \quad (\text{A.71})$$

So one can get the probability that V_k is less than 1 as

$$\begin{aligned} \Pr\{V_k < 1\} &= \frac{1}{\nu^k \gamma(k, \theta)} \int_0^{L_p} (L_p - u)^{k-1} e^{-\frac{L_p}{\nu}} du \\ &= \frac{e^{-\theta}}{\nu^k \gamma(k, \theta)} \frac{L_p^k}{k} \end{aligned} \quad (\text{A.72})$$

Thus, the CDF of V_k^- is

$$\begin{aligned} F_{V_k^-}(x) &= \Pr\{V_k^- \leq x | V_k < 1\} = \frac{\Pr\{V_k < x\}}{\Pr\{V_k < 1\}} \\ &= e^\theta \frac{k}{L_p^k} \int_0^{L_p} (L_p - u)^{k-1} e^{-\frac{1}{\nu}(L_p - u - \frac{u}{x})} du \\ &= \frac{k}{L_p} \int_0^{L_p} \left(\frac{L_p - u}{L_p} \right)^{k-1} e^{\frac{u}{\nu}(1 - \frac{1}{x})} du, \end{aligned} \quad (\text{A.73})$$

for $x \in (0, 1)$ and the average of V_k^- can be expressed as

$$\begin{aligned} \mathbb{E}[V_k^-] &= \int_0^1 x dF_{V_k^-}(x) \stackrel{(a)}{=} x F_{V_k^-}(x) \Big|_0^1 - \int_0^1 F_{V_k^-}(x) dx \\ &= 1 - \frac{k}{L_p} \int_0^1 \int_0^{L_p} \left(\frac{L_p - u}{L_p} \right)^{k-1} e^{\frac{u}{\nu}(1 - \frac{1}{x})} dx du, \end{aligned} \quad (\text{A.74})$$

where (a) follows the integration by parts.

For the case of $k = 0$, define $V_0 = \frac{L_p}{s_n}$ and $V_0^- = V_0 |_{V_0 < 1}$. Specifically, if one packet is completed within one block, the

vestige time equals to its actual service time. The CDF of V_0 and V_0^- are, respectively

$$F_{V_0} = \Pr\{V_0 \leq x\} = \Pr\{s_n \geq \frac{L_p}{x}\} = e^{-\frac{\theta}{x}} \quad (\text{A.75})$$

for $x \in (0, +\infty)$ and

$$F_{V_0^-} = \Pr\{V_0 \leq x | V_0 < 1\} = \frac{\Pr\{V_0 \leq x\}}{\Pr\{V_0 \leq 1\}} = e^{\theta(1 - \frac{1}{x})} \quad (\text{A.76})$$

for $x \in (0, 1)$.

Then we can get the expected value of V_0^- as

$$\begin{aligned} \mathbb{E}[V_0^-] &= \int_0^1 x dF_{V_0^-}(x) \\ &= x e^{\theta(1 - \frac{1}{x})} \Big|_0^1 - \int_0^1 e^{\theta(1 - \frac{1}{x})} dx \\ &= 1 - e^\theta \int_0^1 e^{-\frac{\theta}{x}} dx. \end{aligned} \quad (\text{A.77})$$

Finally, by the whole probability formula, the expected value of vestige time V is

$$\mathbb{E}[V] = \mathbb{E}[V_0^-] \Pr\{T = 0\} + \sum_{k=1}^{\infty} \mathbb{E}[V_k^-] \Pr\{T = k\}, \quad (\text{A.78})$$

where

$$\begin{aligned} &\sum_{k=1}^{\infty} \mathbb{E}[V_k^-] \Pr\{T_n = k\} \\ &= \sum_{k=1}^{\infty} \left[1 - \frac{k}{L_p} \int_0^1 \int_0^{L_p} \left(\frac{L_p - u}{L_p} \right)^{k-1} e^{\frac{u}{\nu}(1 - \frac{1}{x})} dx du \right] \frac{1}{k!} e^{-\theta} \theta^k \\ &= 1 - e^{-\theta} - \frac{k}{L_p} \int_0^1 \int_0^{L_p} \sum_{k=1}^{\infty} \left(\frac{L_p - u}{L_p} \right)^{k-1} \frac{\theta}{k!} e^{-\theta} e^{\frac{u}{\nu}(1 - \frac{1}{x})} dx du \\ &\stackrel{(a)}{=} 1 - e^{-\theta} - \frac{\theta e^{-\theta}}{L_p} \int_0^1 \int_0^{L_p} e^{\theta(1 - \frac{u}{L_p})} e^{\frac{u}{\nu}(1 - \frac{1}{x})} dx du \\ &\stackrel{(b)}{=} 1 - e^{-\theta} - \frac{\theta}{L_p} \int_0^1 \int_0^{L_p} e^{\frac{-u}{\nu x}} dx du \\ &= 1 - e^{-\theta} - \int_0^1 x \left(1 - e^{-\frac{\theta}{x}} \right) dx \\ &= \frac{1}{2} - e^{-\theta} + \int_0^1 x e^{-\frac{\theta}{x}} dx, \end{aligned} \quad (\text{A.79})$$

where (a) follows $\sum_{k=1}^{\infty} \frac{1}{(k-1)!} x^{k-1} = e^x$ and (b) uses the equation $\frac{\theta}{L_p} \cdot \nu = 1$.

Combining (A.77), (A.78) and (A.79) we have

$$\begin{aligned} \mathbb{E}[V] &= e^{-\theta} \cdot \left(1 - e^\theta \int_0^1 e^{-\frac{\theta}{x}} dx \right) + \frac{1}{2} - e^{-\theta} + \int_0^1 x e^{-\frac{\theta}{x}} dx \\ &= \frac{1}{2} + \int_0^1 (x - 1) e^{-\frac{\theta}{x}} dx. \end{aligned} \quad (\text{A.80})$$

Finally, combining (38), (A.66) and (A.80), we have

$$\mathbb{E}[D] = \theta + \frac{\theta^2}{2(1 - \theta)} + \frac{1}{2} + \int_0^1 (x - 1) e^{-\frac{\theta}{x}} dx, \quad (\text{A.81})$$

which completes the proof of Theorem 3. ■

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, p. 379-423/623-656, 1948.
- [2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes," *Proc. 1993 Int. Conf. Comm.*, pp. 1064-1070.
- [3] R. G. Gallager, *Low Density Parity Check Codes*, MIT Press, Cambridge, MA, 1963.
- [4] H. Zimmerman, "OSI reference model-The ISO model of architecture for open systems interconnection", *IEEE Trans. on Commun.*, vol. 28, no. 4, pp. 425-432, Apr. 1980.
- [5] D. Wetteroth, *OSI Reference Model for Telecommunications*, McGraw-Hill, 2001.
- [6] L. Zheng, D. N. C. Tse, and M. Medard, "Channel coherence in the low SNR regime," *IEEE Trans. Inform. Theory*, vol. 53, no. 3, pp. 976-997, March 2007.
- [7] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidot, "Low SNR capacity of noncoherent fading channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1555-1574, April 2009.
- [8] G. Durisi, U. G. Schuster, H. Bolcskei, and S. Shamai (Shitz), "Noncoherent capacity of underspread fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 367-395, Jan. 2010.
- [9] R. McEliece and W. Stark, "Channels with block interference," *IEEE Trans. Inf. Theory*, vol. 30, no. 1, pp. 44-53, Jan. 1984.
- [10] G. Caire and S. Shamai(Shitz), "On the achievable throughput of a multiple- antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691-1706, Jul. 2003.
- [11] A. S. Avestimehr and D. N. C. Tse, "Outage capacity of the fading relay channel in the low SNR regime", *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1401-1415, Apr. 2007.
- [12] L. Li and A. J. Goldsmith, "Capacity and optimal resource allocation for fading broadcast channels: Part II: Outage capacity," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 1103-1127, Mar. 2001.
- [13] M. Zorzi, "Packet dropping statistics of a data-link protocol for wireless local communications", *IEEE Trans. Veh. Technol.*, vol. 52, no. 1, pp. 71-79, Jan. 2003.
- [14] K.K. Lee and S.T. Chanson, "Packet loss probability for bursty wireless real-time traffic through delay model", *IEEE Trans. Veh. Technol.*, vol. 53, no. 3, pp. 929-938, Mar. 2004.
- [15] F. Ishizaki and G. U. Hwang, "Queueing Delay Analysis for Packet Schedulers With/Without Multiuser Diversity Over a Fading Channel", *IEEE Trans. Veh. Technol.*, vol. 56, no. 5, pp. 3220-3227, May, 2007.
- [16] R. Zhang and L. Cai, "Joint AMC and Packet Fragmentation for Error Control Over Fading Channels", *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 3070-3080, June, 2010.
- [17] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630-643, July 2003.
- [18] R. G. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, 1985, vol. 31, no. 2, pp. 124-142.
- [19] I. E. Telatar, R. G. Gallager, "Combining queueing theory with information theory for multiaccess," *IEEE Journal on Selected Areas in Commu.*, 1995, vol. 13, no. 6, pp. 963-969.
- [20] R.A. Berry and R.G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1135-1149, May. 2002.
- [21] Y. Dong, Q. Wang, P. Fan and K. B. Letaief, "The Deterministic Time-Linearity of Service Provided by Fading Channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 5, pp. 1666-1675, May, 2012.
- [22] Y. Dong, P. Fan, K. B. Letaief, and R. D. Murch, " ϵ -Overflow Rate: Buffer-aided Information Transmission over Nakagami- m Fading Channels," in *IEEE 8th International Wireless Commun. & Mobile Comput. Conf.* Limassol, Cyprus August, 2012, pp.19-23.
- [23] Y. Dong, P. Fan, K. B. Letaief and R. D. Murch, "Performance Analysis for Buffer-Aided Communication over Block Rayleigh Fading Channels: Queue Length Distribution, Overflow Probability and epsilon-Overflow Rate", *Wireless Commun. and Mobile Comput.*, vol. 12, no. 18, pp. 1581-1591, Dec, 2012.
- [24] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Veh. Technol.*, vol. 43, no. 2, pp. 359-378, May 1994.
- [25] P. Fan, *Stochastic Processes: Theory and Applications*, Press of Tsinghua University, pp. 84, April. 2006.
- [26] N. Tian, X. Xiu and Z. Ma, *Discrete time queueing theory*, Press of Science, pp. 93, June, 2008.
- [27] Y. Dong and P. Fan, "Queueing Analysis for Block Fading Rayleigh Channels in the Low SNR Regime", in the *5th International Conference on Wireless Communications and Signal Processing*, Hangzhou, China, 2013.
- [28] "Lambert W function," online available at http://en.wikipedia.org/wiki/Lambert-W_function
- [29] Xi'an Jiaotong University Advanced Mathematics Lab, *Complex Variable Functions*, Press of Advanced Education, pp. 87, Aug. 2002.
- [30] F. N. Gouweleew and H. C. Tijms, "Computing loss probabilities in discrete-time queues," *Oper. Res.*, vol. 46, no. 1, pp. 149-154, 1998.
- [31] Yuanguo Zhu, *Matrix analysis and calculation*, National defence industry press, Aug. 2012.
- [32] M. Bhatnagar, A. Hjørungnes and L. Song, Differential coding for non-orthogonal space-time block codes with non-unitary constellations over arbitrarily correlated rayleigh channels, *IEEE Trans. on Wireless Commun.* vol. 8, no. 8, pp. 3985-3995, Aug. 2009.
- [33] L. Song and A. G. Burr, General differential modulation scheme for quasi-orthogonal space-time block codes with partial or full transmit diversity, *IET Commun.*, vol. 1, no.2, pp. 256-266, 2007.